Validation of molecular dynamics simulation

Wilfred F. van Gunsteren
Oxford Centre for Molecular Sciences and New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, England and Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zurich, ETH Zentrum, CH-8092 Zurich, Switzerland

Alan E. Mark
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zurich, ETH Zentrum, CH-8092 Zurich, Switzerland

(Received 21 October 1997; accepted 14 January 1998)

How the results obtained by computer simulation of molecular systems can be validated is considered. The overall quality of the simulated properties of a molecular system will depend on (i) the quality of the theory or model, (ii) the accuracy of the interatomic interaction function or force field, (iii) the degree of sampling, statistics and convergence reached in the simulation, (iv) the quality of the simulation software, and (v) how competently the simulation software is used. These five validation issues are discussed and illustrated with examples. Guidelines for different members of the scientific community are formulated which are aimed at enabling and improving the validation of simulation results in the literature. © *1998 American Institute of Physics.* [S0021-9606(98)50915-4]

I. INTRODUCTION

Computer simulation of molecular systems is playing an ever growing role in academic and industrial research. In areas ranging from materials science and chemistry to pharmacy and molecular biology, computer simulation is already a part of daily practice. Using the molecular dynamics (MD) simulation method the behavior of a variety of molecular systems can be studied. These include liquids, solutions, electrolytes, polymers such as proteins, DNA, and polysaccharides, as well as membranes, liquid crystals, crystals, and zeolites.[1–5] Processes such as melting, adsorption, segregation, formation of molecular complexes, and protein denaturation can be analyzed, and phenomena such as protein stability, enzyme reactivity and membrane permeability can be investigated. Such studies may lead not only to improved understanding and insight, but also to practical results such as engineered proteins or materials with properties optimized for particular applications.

Computer simulation of molecular systems requires software to calculate the interatomic interactions and to integrate the equations of motion. Developing MD simulation software for simple atomic or molecular liquids is a relatively easy task. However, when simulating biomolecular systems with many different types of atoms and interactions, using a variety of boundary conditions, very complex software is required. The software must often manipulate, simulate and analyze thousands or even tens of thousands of atoms. This situation has led to the development of simulation software packages both in academia [AMBER,[6] BRUGEL,[7] CEDAR,[8] CHARMM,[9] EGO,[10] ENCAD,[11] FOCUS,[12] GROMACS,[13] GROMOS,[14] MOIL,[15] NAMD,[16] POLARIS,[17] UHBD,[18] X-PLOR,[19] YASP (Ref. 20)] and by software houses [CHARMm, DISCOVER (Ref. 21), SYBYL (Ref. 22)]. The developers of these molecular simulation software packages are far outnumbered by the users of these packages. The users are generally much less knowledgeable regarding the implemented algorithms and interaction functions than the developers. When using any of these software packages a scientist is inevitably confronted with the question of validation.

A basic task of all scientists is to ensure that the results they obtain can be validated. If the result of a simulation is novel or unexpected or strange, either or both of the following situations may have occurred:

(i) A new phenomenon has been found.
(ii) The results are wrong, because

(1) the model that was used is inappropriate for the application,
(2) the force field is inadequate,
(3) the results have not converged due to insufficient sampling,
(4) the software contains bugs,
(5) the software has been used incorrectly.

Before concluding an exciting new phenomenon has been found, situation (ii) must be ruled out. The reply to each of these five validation questions must be no. However, it is only possible to answer these questions if,

(1) a full description of the model and algorithms is readily available,
(2) a full description of the interaction function or force field is readily available,
(3) simulation results are shown as a function of simulation length,
(4) the source code of the software can be checked,
(5) the set-up of the simulations is described in detail.

Unfortunately, these conditions are often far from being fulfilled. Software manuals tend to describe input parameters

rather than algorithms. In many cases source code is not accessible, and force fields are generally not completely described in the literature. In regard to force fields, note that it is not sufficient to quote the functional form of the interaction function and to give lists of parameters.[11,24,25] The assignment of dihedral angle types to actual proper and improper torsional angles in molecules, the treatment of first, second, and third covalently bound neighbors and the use of particular combination rules when calculating nonbonded interactions, factors rarely mentioned in the literature, should be also specified, for example.[14,26]

A validation of molecular models and force fields generally involves a comparison of simulated with experimental data. A proper interpretation of the latter is therefore essential to validation. This is briefly discussed in the next section.

Every scientist who uses simulation software will be repeatedly confronted with the five mentioned validation questions. In the next section, we will elaborate on and illustrate these five issues using examples taken from our own work and the literature. The examples are only briefly described, the details being left to the referenced literature. Literature on how to perform computer simulations can be found in Refs. 1–5. In the last section of this paper we discuss what might be done by members of the scientific community to enable better validation of simulation results and reports by users of simulation software and readers of the literature.

II. FIVE VALIDATION ISSUES

When attempting to validate the results of a molecular simulation, the following issues should be considered:

(1) *The quality of the theory or model.* The choice of molecular, atomic or electronic degrees of freedom that are explicitly simulated, the type of equations of motion used and the treatment of the boundary of the system will determine the results that can be obtained. The quality of the assumptions and approximations inherent to the molecular model used will determine the accuracy of the simulated results.

(2) *The accuracy of the interatomic interaction function or force field.* The choice of functional form, the parameter values, the theoretical or experimental data used to calibrate the interaction function and the calibration procedure will determine the accuracy of a force field. The accuracy of a force field may vary with the type of molecule, the phase (gas vs liquid) and type of property that is considered.

(3) *The degree of sampling, statistics, and convergence reached in the simulation.* The simulation period should be much longer than the relaxation time of the property considered. The relaxation time of a property will depend on the type of property, the thermodynamic state (temperature, pressure), and the type of molecule.

(4) *The quality of the simulation software.* The more complex software becomes, the more difficult it is for software developers to ensure its correctness.

(5) *How competently the simulation software is used.* When using complex software with a multitude of input parameters, a parameter combination that induces erroneous results is easily selected.

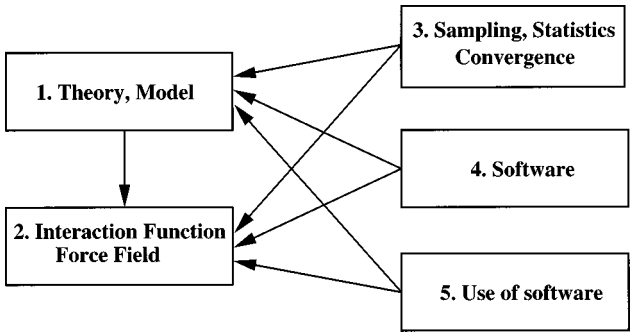

FIG. 1. Five issues regarding the validation of results of molecular simulation. The arrows indicate whether the quality of the handling of one issue will influence the test of the quality of the issue the arrow is pointing at.

When performing tests with respect to the five mentioned validation issues it should be kept in mind that the issues depend on each other as is illustrated by the arrows in Fig. 1. For example, a force field test makes only sense when involving converged properties simulated using bug-free software with appropriate input parameters and an adequate molecular model. Or, convergence characteristics of a given molecular property will generally not depend on the details of the interaction function.

Simulation studies are normally verified by a comparison of simulated and experimentally measured properties of the system considered. The results of such a comparison between simulation and experiment can be classified as follows:

(A) *Agreement between simulation and experiment is obtained.*

This may be due to one or more of the following reasons:

(1) The simulation adequately reflects the experimental system.
(2) The property examined is insensitive to the details of the simulation.
(3) A compensation of errors has occurred.

(B) *No agreement between simulation and experiment is obtained.*

This may be due to one or both of the following reasons:

(1) The simulation does not reflect the experimental system. Either, the theory or model is incorrect, the force field used is inadequate, the simulation is not converged, the software is at fault, or the software is incorrectly used.
(2) The experimental data are incorrect.

An example of a property that can be insensitive to the chosen simulation parameters (case A2) is given in Fig. 2. This shows the agreement of the J-coupling constants of a β-peptide simulated at two different temperatures with the experimental values. Although the fold of the peptide is helical at room temperature and extended at high temperature, this large structural difference is not mirrored in the simulated J-values; the agreement with the experimental J-values is

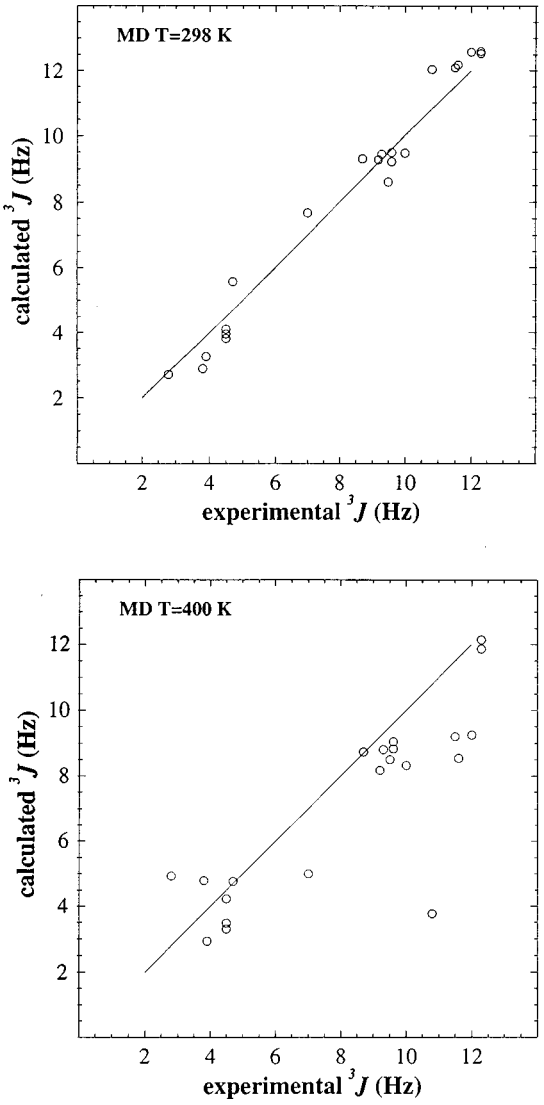

FIG. 2. Simulated vs experimental ^3J-coupling constants for a β-heptapeptide in methanol. The simulated values are averages over 2 ns MD simulations at room temperature (298 K) and at elevated temperature (400 K). Data taken from Ref. 27.

rather insensitive to the particular fold of the heptapeptide.[27] Examples of observed compensation of errors (case A3) have been collected in the literature.[28] When testing simulation results by comparison with experimental data it should be remembered that good agreement between simulated and experimentally measured properties does not necessarily imply that the simulation is correct. The good agreement may simply result from the compensation of errors.

A. First validation issue: Quality of theory or model

In this section we give a few examples of inadequate theory or modelling influencing the results of a molecular simulation.

Electrostatic interactions are extremely long range in nature and approximations in their treatment are therefore necessary. For nonpolar or nonionic systems neglect of nonbonded interactions beyond a long cutoff distance, e.g., 1.5 nm, may be a reasonably accurate approximation. For simulating ionic systems this model is totally inadequate, as is

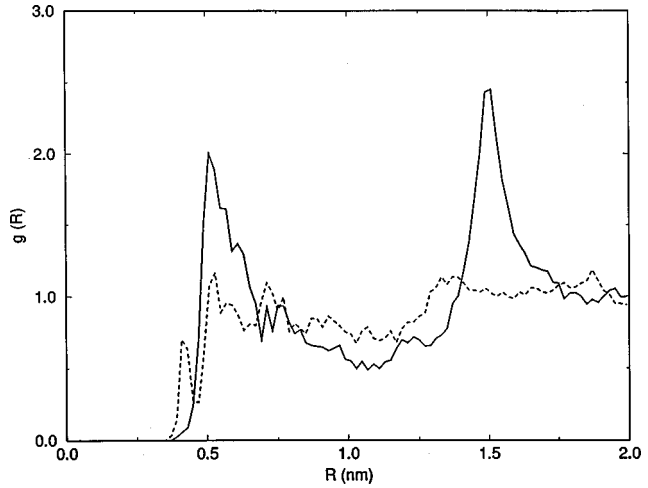

FIG. 3. Radial distribution function $g(R)$ between chlorine ions calculated from MD simulations of a periodic box containing 40 Na^+, 40 Cl^-, and 2127 water molecules using different treatments of long-range electrostatic forces; spherical truncation at $R=1.5$ nm (solid line) and using a reaction field (dashed line). Data taken from Ref. 29.

demonstrated in Fig. 3, which shows the radial distribution function for chloride–chloride ion pairs obtained from a MD simulation of 40 Na^+ and 40 Cl^- ions solvated in a periodic box with 2127 water molecules, a 1.0 M NaCl solution.[29] The box was quite large, with an edge length of 4.05 nm, and so was the cutoff radius of 1.5 nm. The effect of direct truncation is to preferentially arrange the ions just outside the cutoff spheres (solid line). When a nonbonded interaction model including a reaction field force due to the dielectric medium out side the cutoff sphere is used,[30] the artefacts at the cutoff distance have disappeared within 50 ps of simulation (dashed line). But, the short-range structure has also changed considerably. So, a (truncation) model that is of reasonable quality for simulations of nonionic systems, can be totally inadequate for ionic ones.

The second example of incorrect theory or modelling concerns the decomposition of the free energy change

$$\Delta F \equiv F(B) - F(A) \tag{1}$$

corresponding to a change of the Hamiltonian of the system

$$\Delta H \equiv H(B) - H(A) \tag{2}$$

from a state A to a state B into two or more components, e.g., ΔF_1 and ΔF_2, that are defined spatially or with respect to individual terms of the interaction function.[31,32] We assume that the Hamiltonian $H(X)$, with $X=A$ or B, can be expressed as a sum of two terms,

$$H(X) = H_1(X) + H_2(X), \tag{3}$$

where the indices 1 and 2 refer, e.g., to spatially different parts of the molecular system X (amino acid residues in proteins, solvent) or to different types of interaction terms in the force field (covalent, van der Waals, electrostatic interactions). Using statistical mechanics one can write

$$\Delta F = -\beta^{-1} \ln[\langle \exp(-\beta \Delta H) \rangle_A], \tag{4}$$

where $\langle \cdots \rangle_A$ denotes averaging over a canonical ensemble generated using the Hamiltonian $H(A)$ and β^{-1} is the prod-

uct of the Boltzmann constant k_B and the temperature T. Using the decomposition (3) in (4), expanding the two exponential functions in Taylor series in ΔH_1 and ΔH_2 and the function $\ln(1-x)$ in a Taylor series in x, and keeping terms up to first order in β, one finds

$$\begin{aligned}\Delta F &= \langle \Delta H_1 \rangle_A - (\beta/2)[\langle (\Delta H_1)^2 \rangle_A - \langle \Delta H_1 \rangle_A^2] + O(\beta^2) \\ &\quad + \langle \Delta H_2 \rangle_A - (\beta/2)[\langle (\Delta H_2)^2 \rangle_A - \langle \Delta H_2 \rangle_A^2] + O(\beta^2) \\ &\quad - \beta[\langle \Delta H_1 \Delta H_2 \rangle_A - \langle \Delta H_1 \rangle_A \langle \Delta H_2 \rangle_A] + O(\beta^2) \\ &= \Delta F_1 + \Delta F_2 - \beta[\langle \Delta H_1 \Delta H_2 \rangle_A - \langle \Delta H_1 \rangle_A \langle \Delta H_2 \rangle_A] \\ &\quad + O(\beta^2), \end{aligned} \tag{5}$$

which shows that the decomposition of ΔF into components ΔF_1 and ΔF_2 is incorrect,

$$\Delta F \neq \Delta F_1 + \Delta F_2, \tag{6}$$

unless ΔH_1 and ΔH_2 are uncorrelated. The cross-correlations between ΔH_1 and ΔH_2 have been calculated for a variety of decompositions (3) with respect to the change of solvation free energy of a set of para-substituted phenols in water.[32] Correlations up to 50% were obtained, indicating that the cross-correlations between parts of a molecular system or between force field terms are generally not negligible for realistic simulations. This implies that the theory of decomposition of free energy differences into components appears to be invalid for many systems of interest.[31,32]

A third, rather subtle example has to do with an incorrect formulation of Hamiltonian mechanics.[33] The Hamiltonian of a system must be expressed as a function of the generalized coordinates $\mathbf{q}$ and their conjugate momenta $\mathbf{p}$, e.g.,

$$H(\mathbf{p},\mathbf{q}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{q}). \tag{7}$$

The first term represents the kinetic energy of a particle with mass m. Using thermodynamic integration the free energy change as a function of the mass m is

$$\Delta F \equiv F(m_B) - F(m_A) = \int_{m_A}^{m_B} \left\langle \frac{\partial H}{\partial m} \right\rangle_m dm \tag{8}$$

or using expression (7)

$$\Delta F = -\int_{m_A}^{m_B} \left\langle \frac{\mathbf{p}^2}{2m^2} \right\rangle_m dm. \tag{9}$$

If the Hamiltonian is expressed as function of the generalized coordinates $\mathbf{q}$ and the associated velocities $\mathbf{v}=\mathbf{p}/m$, one has

$$H(m\mathbf{v},\mathbf{q}) = \tfrac{1}{2} m\mathbf{v}^2 + V(\mathbf{q}) \tag{10}$$

which leads, using Eq. (8), to

$$\Delta F = +\int_{m_A}^{m_B} \left\langle \frac{\mathbf{v}^2}{2} \right\rangle_m dm = +\int_{m_A}^{m_B} \left\langle \frac{\mathbf{p}^2}{2m^2} \right\rangle_m dm. \tag{11}$$

This expression has the wrong sign due to the use of the velocity $\mathbf{v}$ as variable in the Hamiltonian, which is not allowed.[33]

A fourth example of inadequate theory or modeling is in regard to procedures to determine biomolecular structure based on nuclear magnetic resonance (NMR) spectroscopic or x-ray diffraction data. The standard procedure is to vary the structure of a single molecule with the aim of obtaining a molecular structure that reproduces as good as possible the experimental data (NOE intensities or distances, J-coupling constants, crystallographic structure factor amplitudes).[34] However, the experimental data represent an average over time and space (molecules). Fitting a single molecular structure to averaged data can lead to very wrong molecular structures, as has been amply demonstrated in the literature.[34–36]

B. Second validation issue: Accuracy of interatomic interaction function or force field

The validation of a force field should involve as many different properties for different types of molecules and environments as possible. For molecular systems three general types of properties can be distinguished.

(1) *Structural properties* (including first or second moments of distributions of properties that depend on molecular configuration) such as

(a) average atom positions or atom-atom distances,
(b) radius of gyration,
(c) solvent accessible surface area,
(d) NMR order parameters (S^2),
(e) crystallographic temperature factors,
(f) dipole moment fluctuations ($\mathbf{M}^2$) leading to an estimate of the dielectric permittivity ϵ,
(g) radial distribution functions ($g(r)$),
(h) density.

(2) *Energetic properties* such as

(a) heat of vaporization,
(b) free energy of solvation,
(c) heat capacity,
(d) isothermal compressibility,
(e) thermal expansion coefficient,
(f) surface tension.

(3) *Dynamical properties* such as

(a) diffusion constants,
(b) rotational correlation times,
(c) dielectric correlation times,
(d) viscosity.

We note that the different properties possess very different relaxation times, which means that MD simulations of very different lengths are required to obtain converged simulated values suitable for comparison to test or calibration data.[37] For example, the radial distribution functions and molecular translational and rotational diffusion coefficients of liquid dimethyl sulfoxide (DMSO) can be obtained from 50 ps simulation of 432 DMSO in a periodic box, whereas the determination of collective properties such as the dielectric permittivity or shear viscosity requires a simulation length of over 1000 ps for this system.[38]

An example of a comparison of structural properties of the 129 residue protein hen egg white lysozyme (HEWL) simulated using three different force fields is shown in Fig. 4.[39] In the simulation in vacuo using the corresponding force

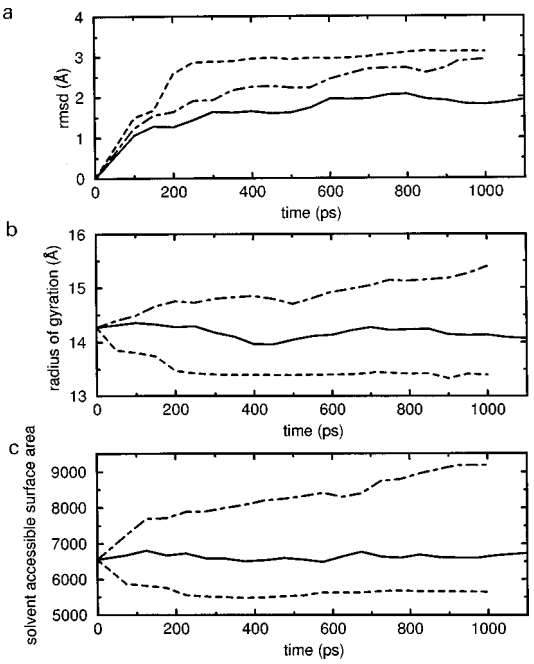

FIG. 4. Structural properties of hen egg white lysozyme (HEWL) as a function of MD simulation time using different force fields; vacuum simulation and corresponding GROMOS force field (dashed line), simulation in water using the GROMOS87 force field (dot–dashed line) and using the GROMOS96 force field (solid line); (a) root-mean-square positional difference for the C_α atoms from the crystal starting structure (Å), (b) radius of gyration (Å), (c) solvent accessible surface area (Å). Data taken from Ref. 39.

field (dashed line) the structure becomes too compact and deviates rapidly from the starting crystal structure. The simulation in a periodic box with a few thousand water molecules using the GROMOS87 force field[40] (dot–dashed line) shows a slow gradual expansion of the molecule on a time scale of hundreds of picoseconds. This implies that the (artefactual) driving forces for this expansion are very small. In fact they were not observed in another equally long simulation of the protein BPTI.[41] The simulation involving the GROMOS96 force field[14] (solid line) produces significantly better structural properties. Figure 4 also illustrates that the relaxation time of structural properties of proteins lies in the range upwards from 100 ps.

An example of a comparison of energetic properties is given in Table I.[42] The relative energy and free energy of binding of para-substituted phenols as guest molecules to α-cyclodextrin as host molecule were calculated from MD simulations of the guest molecules and of the complexes in aqueous solution. Entropic contributions to the binding seem to play an important role for these flexible molecules. From thermodynamic cycle closure the lower error bound for the ΔG_{calc} values was estimated to be about $k_B T = 2.5$ kJ mol^{-1}. The root-mean-square deviation with the experimental values is of the same order of magnitude, which implies that simulation and experiment agree within the error bound set by the extent of sampling in the simulations.

Finally, we must stress that force field validation can only be carried out using equilibrated systems and comparing converged average values of properties.

TABLE I. Relative energy (ΔE_{calc}) and free energy (ΔG_{calc}) of binding of para-substituted phenols to α-cyclodextrin in aqueous solution. The ΔE_{calc} values are the averaged host–guest interaction energies, whereas the ΔG_{calc} values are obtained from free energy perturbation simulations in which the guest molecules were mutated when bound to the host molecule (in a periodic box with 508 water molecules) and when unbound (in a periodic box with 544 water molecules). Data taken from Ref. 42.

	ΔE_{calc}	ΔG_{calc} kJ mol^{-1}	ΔG_{exp}
p-methylphenol	0.0	0.0	0.0
p-chlorophenol	−11.2	−8.1	−3.9
p-cyanophenol	−23.6	−5.9	−2.9
p-methoxyphenol	−13.7	−4.2	−0.1
root mean square deviation from ΔG_{exp}	14.9	3.8	0.0

C. Third validation issue: Degree of sampling, statistics, and convergence

The central question regarding the third validation issue is whether the length of a MD simulation is sufficiently long to yield reliable trajectory averages of the different molecular or system properties of interest. Trajectory averages will generally only be representative when the equilibration time of the simulation, τ_{equil}, is longer than the relaxation time $\tau_{\mathrm{relax}}(Q)$ of the property Q,

$$\tau_{\mathrm{equil}} > \tau_{\mathrm{relax}}(Q), \tag{12}$$

and when the sampling time, τ_{sample}, is much longer than $\tau_{\mathrm{relax}}(Q)$,

$$\tau_{\mathrm{sample}} \gg \tau_{\mathrm{relax}}(Q). \tag{13}$$

If conditions (12) and (13) are not fulfilled, the trajectory average $\langle Q(t)\rangle$ of the property Q will display a drift as a function of time or erratic behaviour due to the occurrence of rare events affecting the value of $Q(t)$.

The relaxation time $\tau_{\mathrm{relax}}(Q)$ may be long for different reasons.

(1) The system may jump rapidly but rarely between relatively stable states. An example is the 180° flipping of phenylalanine side chains in a protein which is a fast, picosecond time-scale process occurring comparatively infrequently, only on a microsecond time scale. In such a case the trajectory averages will be sensitive to the number of rare events that are simulated.
(2) The system may change intrinsically slow, in which case trajectory averages will display a continuous change as a function of time.

The relaxation and dynamics of the different properties occurring in a MD simulation can be analyzed by different means.

(1) For equilibration simulations one may monitor the time series of a property $Q(t)$, or of its average $\langle Q(t)\rangle$ or fluctuations $\langle [Q(t)-\langle Q(t)\rangle]^2\rangle^{1/2}$, or calculate its autocorrelation function $\langle Q(t')Q(t'+t)\rangle$. The decay time of the autocorrelation function or the build-up rates of the trajectory averages give an indication of $\tau_{\mathrm{relax}}(Q)$.

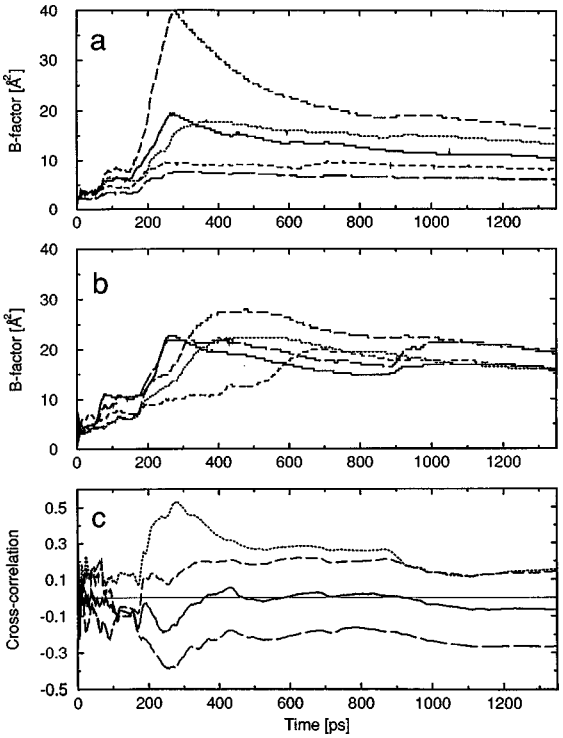

FIG. 5. Time development of the mean square positional fluctuation or *B*-factor (14) for a number of C_α atoms in the protein BPTI as calculated from a 1.4 ns MD simulation of BPTI in a periodic box with 2371 water molecules. (a) For α-helix residues 50 (short dash), 51 (long dash), 53 (solid), 54 (medium dash), and 55 (dotted), (b) for 3_{10}-helix residues 3 (solid), 4 (long dash), 5 (medium dash), 6 (dotted), and 7 (short dash). (c) Cross-correlation coefficients between the motions of C_α atom pairs surrounding the 5–55 disulfide bridge in BPTI, atom pairs 5–52 (long dash), 5–53 (solid), 5–54 (medium dash), and 5–55 (dotted). Data taken from Ref. 43.

(2) When starting a simulation from a non-equilibrium initial state, the rate of relaxation towards equilibrium for different properties $Q(t)$ will give an indication of $\tau_{\mathrm{relax}}(Q)$.
(3) If different MD simulations starting from different initial states do not converge to the same trajectory average for property Q, it can be concluded that $\tau_{\mathrm{relax}}(Q)$ is longer than the simulation time.

An overview of the relaxation behavior and convergence characteristics of trajectory averages for a variety of properties of proteins in aqueous solution has been presented in Ref. 37. Energetic and structural quantities, NMR relaxation parameters, dielectric relaxation, free energy of complexation, solvent, and ion dynamics were considered. The examples shown in Ref. 37 will not be repeated here. Only the slow convergence of the atomic positional fluctuations for a protein in aqueous solution is illustrated in Fig. 5. It shows the so-called *B*-factor,

$$B_i=(8\pi^2/3)(\langle \mathbf{r}_i^2\rangle-\langle \mathbf{r}_i\rangle^2), \tag{14}$$

which is proportional to the mean square positional fluctuation, for a number of atoms i in the protein BPTI as a function of averaging time.[43] Even for atoms in relatively stable helical parts of the protein the positional fluctuations are only beginning to converge after hundreds of picoseconds. The cross-correlations between neighbouring atoms display very erratic behavior when the sampling is less than about 200 ps. These results and the cases presented in Ref. 37 illustrate that trajectory averages obtained from simulations should be analyzed and interpreted with a clear eye to the limitations of sampling on the simulation time scale.

D. Fourth validation issue: Quality of the simulation software

The quality of simulation software depends primarily on the care with which it was constructed and tested by the software developers. Testing of simulation software can be done on various levels.

(1) Elementary algorithmic tests for MD simulation codes involve elementary classical mechanical laws.

(a) It can be numerically tested for each term of the force field that the force on each atom is equal to the negative numerical gradient of the energy.
(b) The total energy of a system without external forces or coupling to temperature or pressure baths should remain constant during a simulation. In other words, the root mean square fluctuation of the total energy, $E_{\mathrm{tot}}=E_{\mathrm{kin}}+E_{\mathrm{pot}}$, of the system,
$$\Delta E_{\mathrm{tot}}=\langle[E_{\mathrm{tot}}-\langle E_{\mathrm{tot}}\rangle]^2\rangle^{1/2} \tag{15}$$
should be small compared to the root mean square fluctuation of the kinetic energy, ΔE_{kin}, or of the potential energy, ΔE_{pot},[44]
$$\Delta E_{\mathrm{tot}}\ll\Delta E_{\mathrm{kin}} \text{ or } \Delta E_{\mathrm{pot}}. \tag{16}$$
We note that the relative fluctuation of the total energy,
$$\Delta E_{\mathrm{tot}}/\langle E_{\mathrm{tot}}\rangle \tag{17}$$
is not a useful criterion for energy conservation, because E_{tot} is only defined up to a constant, which implies that (17) will be arbitrarily large or small depending on the choice of the origin of the energy scale.
(c) The total momentum of the system should remain constant in the absence of external forces or coupling to temperature or pressure baths.

(2) The simulation software is used to reproduce well-known standard results for benchmark systems from the literature.

(3) The simulation software is used in practical research projects.

(4) By making the software including the source code available to the scientific community, the extent of de facto testing is greatly enhanced.

Users of standard simulation software packages could test the quality of the obtained software themselves by executing the types of tests mentioned above. Another possibility is to compare results obtained using different simulation software packages. This option is not easily executed in practice, since it requires the same algorithms and force fields to be present in both simulation software packages.

TABLE II. Guidelines for different members of the scientific community aimed at enabling and improvement of validation of simulation results.

Validation issue	(A) Software suppliers/vendors	(B) Software users	(C) Publishers, editors reviewers
(1) Theory, model	Specify: theory, equations	Specify (or reference): theory, equations	Require: author compliance
(2) Force field	Specify: complete form, parameters, parameter assignments, version code	Specify: version code modifications	Require: author compliance
(3) Sampling, convergence		Show: time evolution of key properties	Require: author compliance
(4) Simulation software	Provide: source code (or proof of correctness) version number	Specify: version number modifications	Require: author compliance
(5) Use of software		Specify: key input parameters	Require: author compliance

E. Fifth validation issue: How competently the simulation software is used

Any software can be used such that nonsensical results are produced. Incompatible values for input parameters can be chosen or parameter values may violate the range of applicability of the molecular model or algorithm. A multifunctional simulation code for biomolecular systems will contain, apart from the many force field parameters, many input parameters to be set by the user. For example, the simulation code of the GROMOS96 simulation package has more than 100 input variables which may adopt two or more values. This situation makes it impossible to bar a determined user from using nonsensical input parameter values.

Most errors can, however, be avoided by a careful consideration of the physical laws involved and the physical and chemical characteristics of the molecular system of interest. For example, if the pressure of the system, $P(t)$, is calculated using the virial formula,[14]

$$P(t)=\tfrac{2}{3}[E_{\mathrm{kin}}(t)-W(t)]/V(t), \tag{18}$$

the pressure will depend on the kinetic energy, $E_{\mathrm{kin}}(t)$, the virial $W(t)$, and the volume $V(t)$ of the periodic box. When coupling the system to a temperature bath and to a pressure bath, the coupling to the former should be tighter than to the latter in order to avoid resonance of $P(t)$ and $E_{\mathrm{kin}}(t)$ induced by Eq. (18). The MD integration time step should be chosen much smaller than the shortest oscillation period present in the system.

When reporting simulation results in the literature simulation parameters should be reported in order to enable readers to judge the setup of a simulation.

III. ENABLING AND IMPROVING VALIDATION OF SIMULATION RESULTS

Having reviewed various aspects of validation of results of MD simulations of molecular systems, the question that remains is what can be done by the various members of the scientific community to enable and improve validation of MD simulation. Answering this question different categories of actors should be distinguished, see Table II.

(A) Software suppliers and vendors.
(B) Software users.
(C) Publishers, editors, reviewers of scientific papers.

What can these different actors do with respect to the validation issues (1)–(5) mentioned in the Introduction? The answers are summarized in Table II.

(A) *Software suppliers and vendors* should allow for proper validation of the software by the following actions:

(i) specify the implemented models and algorithms (issue 1),
(ii) specify completely the force fields provided, including functional form, parameters, assignments, constants, and give different force field versions different version codes for identification (issue 2),
(iii) provide source code and version number for checking and identification, or if this is impossible, provide standard results for checking (issue 4).

(B) *Software users* should allow for proper validation of their simulation results by the following actions:

(i) refer to the theory or model that was used in the simulation (issue 1),
(ii) refer to the force field used giving the version code and add information on modifications that were made (issue 2),
(iii) present the time evolution of key properties of the molecular system for judging the degree of sampling and convergence (issue 3),
(iv) refer to the version number of the software that was used and specify the modifications made (issue 4),
(v) specify the chosen input parameters of the simulation (issue 5).

(C) *Publishers, editors, and reviewers of scientific papers* should enforce these guidelines for simulation software suppliers and users.

If the simulation community adheres to such guidelines, it will become much easier for a scientist applying simulation methods to validate the obtained simulation results and to convince him or herself that the novel or unexpected or strange results are not due to a flawed model, an inadequate force field, insufficient sampling, software bugs or input errors, but instead are indicative of a new phenomenon.

ACKNOWLEDGMENT

W. F. van Gunsteren gratefully acknowledges financial support given by the Underwood Fund.

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[2] *Computer Modeling of Fluids, Polymers and Solids*, NATO ASI Series C293, edited by C. R. A. Catlow, S. C. Parker, and M. P. Allen (Kluwer, Dordrecht, 1990).
[3] *Computer Simulation in Material Science*, NATO ASI Series E205, edited by M. Meyer and V. Pontikis (Kluwer, Dordrecht, 1991).
[4] *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson (ESCOM, Leiden, 1993), Vol. 2.
[5] *Quantum Mechanical Simulation Methods for Studying Biological Systems*, edited by D. Bicout and M. Field (Springer, Berlin, 1996).
[6] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, Comput. Phys. Commun. **91**, 1 (1995).
[7] P. Delhaise, D. van Belle, M. Bardiaux, A. Alard, P. Hamers, E. van Cutsem, and S. J. Wodak, J. Mol. Graphics **3**, 116 (1985).
[8] M. Carson and J. Hermans, in *Molecular Dynamics and Protein Structure*, edited by J. Hermans (University of North Carolina, Chapel Hill, 1985), pp. 165–166.
[9] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[10] M. Eichinger, H. Grubmüller, and H. Heller, *User Manual for EGO-VIII, Release 1.0* (Universität München, München, 1995).
[11] M. Levitt, M. Hirshberg, R. Sharon, and V. Daggett, Comput. Phys. Commun. **91**, 215 (1995).
[12] A. P. Lemon, P. Dauber-Osguthorpe, and D. J. Osguthorpe, Comput. Phys. Commun. **91**, 97 (1995).
[13] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[14] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The* GROMOS96 *Manual and User Guide* (Hochschulverlag an der ETH Zürich, Zürich, 1996).
[15] R. Elber, A. Roitberg, C. Simmerling, R. Goldstein, H. Li, G. Verkhivker, C. Keasar, J. Zhang, and A. Ulitsky, Comput. Phys. Commun. **91**, 159 (1995).
[16] M. Nelson, W. Humphrey, R. Kufrin, A. Gursoy, A. Dalke, L. Kale, R. Skeel, and K. Schulten, Comput. Phys. Commun. **91**, 111 (1995).
[17] F. S. Lee, Z. T. Chu, and A. Warshel, J. Comput. Chem. **14**, 161 (1993).
[18] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon, Comput. Phys. Commun. **91**, 57 (1995).
[19] A. T. Brünger, *X-PLOR: A System for X-ray Crystallography and NMR* (Howard Hughes Medical Institute and Yale University, New Haven, 1996).
[20] F. Müller-Plathe, Comput. Phys. Commun. **78**, 77 (1993).
[21] Molecular Simulations Inc., Waltham, MA 02154, 1997.
[22] Tripos Inc., St. Louis, MO 63144, 1997.
[23] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. Spellmayer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[24] A. D. MacKerrel, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, D. T. Nguyen, T. Ngo, B. Prodham, B. Roux, B. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus, Biophys. J. **61**, A143 (1992).
[25] M. J. Hwang, T. P. Stockfish, and A. T. Hagler, J. Am. Chem. Soc. **116**, 2515 (1994).
[26] W. F. van Gunsteren and M. Karplus, Macromolecules **15**, 1528 (1982).
[27] X. Daura, W. F. van Gunsteren, D. Rigo, B. Jaun, and D. Seebach, Chem. Eur. J. **3**, 1410 (1997).
[28] W. F. van Gunsteren, in *Studies in Physical and Theoretical Chemistry*, edited by J.-L. Rivail (Elsevier, Amsterdam, 1990), Vol. 71, pp. 463–478.
[29] P. E. Smith and W. F. van Gunsteren, in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications* (ESCOM, Leiden, 1993), Vol. 2, pp. 182–212.
[30] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, J. Chem. Phys. **102**, 5451 (1995).
[31] A. E. Mark and W. F. van Gunsteren, J. Mol. Biol. **240**, 167 (1994).
[32] P. E. Smith and W. F. van Gunsteren, J. Phys. Chem. **98**, 13735 (1994).
[33] W. F. van Gunsteren, in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications* (ESCOM, Leiden, 1989), pp. 27–59.
[34] W. F. van Gunsteren, R. M. Brunne, P. Gros, R. C. van Schaik, C. A. Schiffer, and A. E. Torda, Meth. Enzym. **239**, 619 (1994).
[35] C. A. Schiffer, P. Gros, and W. F. van Gunsteren, Acta Crystallogr., Sect. D: Biol. Crystallogr. **51**, 85 (1995).
[36] A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, J. Mol. Biol. **214**, 223 (1990).
[37] W. F. van Gunsteren, P. H. Hünenberger, A. E. Mark, P. E. Smith, and I. G. Tironi, Comput. Phys. Commun. **91**, 305 (1995).
[38] H. Liu, F. Müller-Plathe, and W. F. van Gunsteren, J. Am. Chem. Soc. **117**, 4363 (1995).
[39] L. J. Smith, A. E. Mark, C. M. Dobson, and W. F. van Gunsteren, Biochemistry **34**, 10918 (1995).
[40] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation* (GROMOS) *Library Manual* (Biomos, University of Groningen, Groningen, 1987).
[41] R. M. Brunne, K. D. Berndt, P. Güntert, K. Wüthrich, and W. F. van Gunsteren, Proteins **23**, 49 (1995).
[42] A. E. Mark, S. P. van Helden, P. E. Smith, L. H. M. Janssen, and W. F. van Gunsteren, J. Am. Chem. Soc. **116**, 6293 (1994).
[43] P. H. Hünenberger, A. E. Mark, and W. F. van Gunsteren, J. Mol. Biol. **252**, 492 (1995).
[44] W. F. van Gunsteren and H. J. C. Berendsen, Mol. Phys. **34**, 1311 (1977).